EDC™ Environmental Data Connector

an extension for connecting ArcGIS to Thredds/OPeNDAP

User's Guide

INTRODUCTION

The Environmental Data Connector (EDC) is a tool that allows users to connect to THREDDS/OPeNDAP servers from within ArcGIS and download environmental data. It was developed for the National Oceanic and Atmospheric Administration (NOAA) Fisheries by Applied Science Associates, Inc. (ASA). The EDC is available in two versions: the original EDC which is designed to work in conjunction with ArcGIS, and the EDC Standalone version which can be run outside ArcGIS.

INSTALLATION

The Environmental Data Connector (EDC) is available for download at

http://www.pfeg.noaa.gov/products/EDC

For the ArcGIS version of the EDC, click the 'Download EDC' link. For the Standalone version of the EDC, click the 'Download EDC Standalone' link.

Although registration is not necessary to download the EDC, we encourage you to register on this website if you would like to be informed when updates are made available. Installation instructions and this user guide are also available at this website.

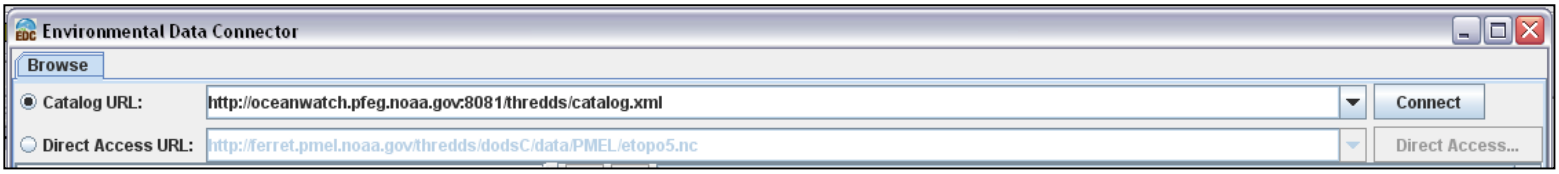
GETTING STARTED

An installation manual is also available for download from the website, Install the EDC according to those instructions.

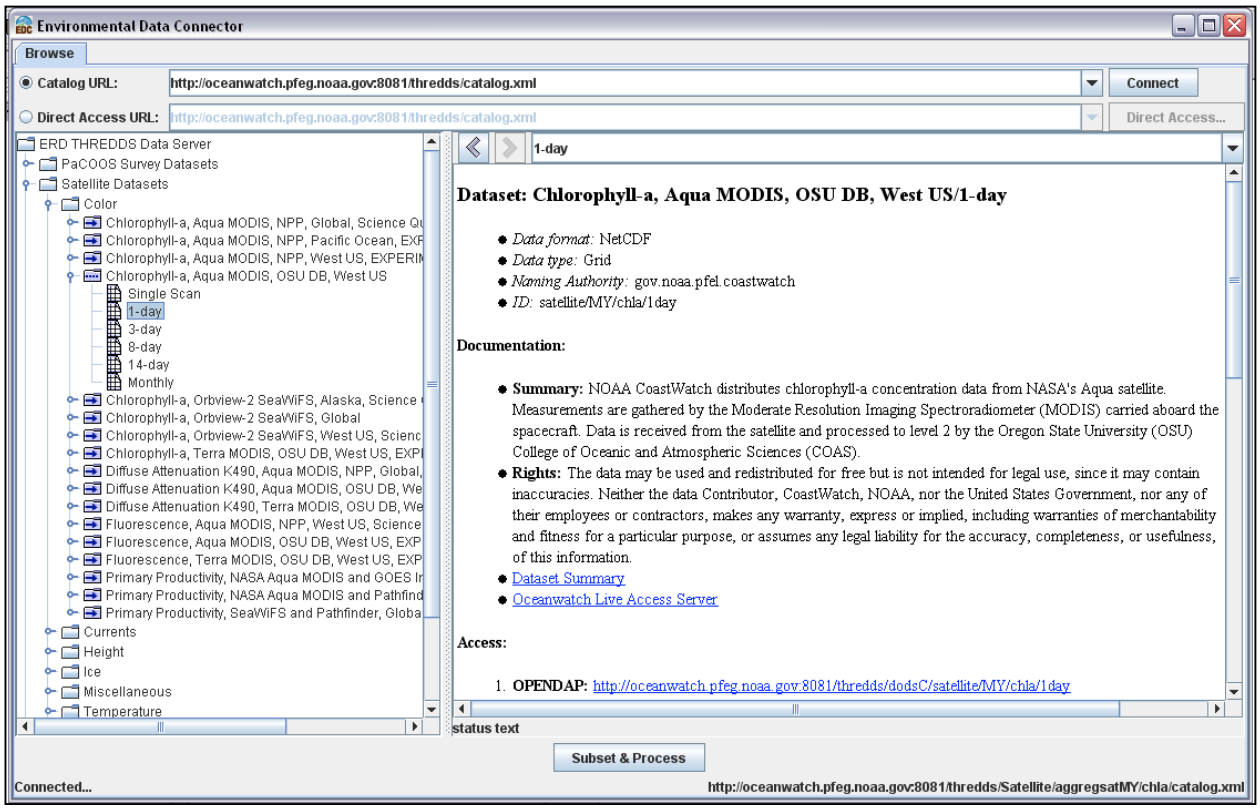
To start the EDC, either click on the EDC icon in ArcMap or click the 'EDC' menu and choose 'Environmental Data Connector'. The EDC will open in a new window. While the EDC is open, ArcGIS is locked and inaccessible.

CONNECTING TO DATA

In the top portion of the EDC, you may select the data source. Valid data sources are either THREDDS data catalogs or individual datasets. The dropdown menus list some common sources for oceanographic and atmospheric data. You may select one of the catalogs or datasets in the dropdown menus or manually type in a different URL. The 'Catalog URL' connects to THREDDS dataset catalogs, while the 'Direct Access URL' connects to individual datasets.



Once you have selected a catalog (or dataset), click the 'Connect' button to connect to that dataset. If you connect to a data catalog, a directory tree will appear in the left hand side of the EDC showing the directory structure of the data catalog. Select the dataset you would like to download by navigating through this directory structure.

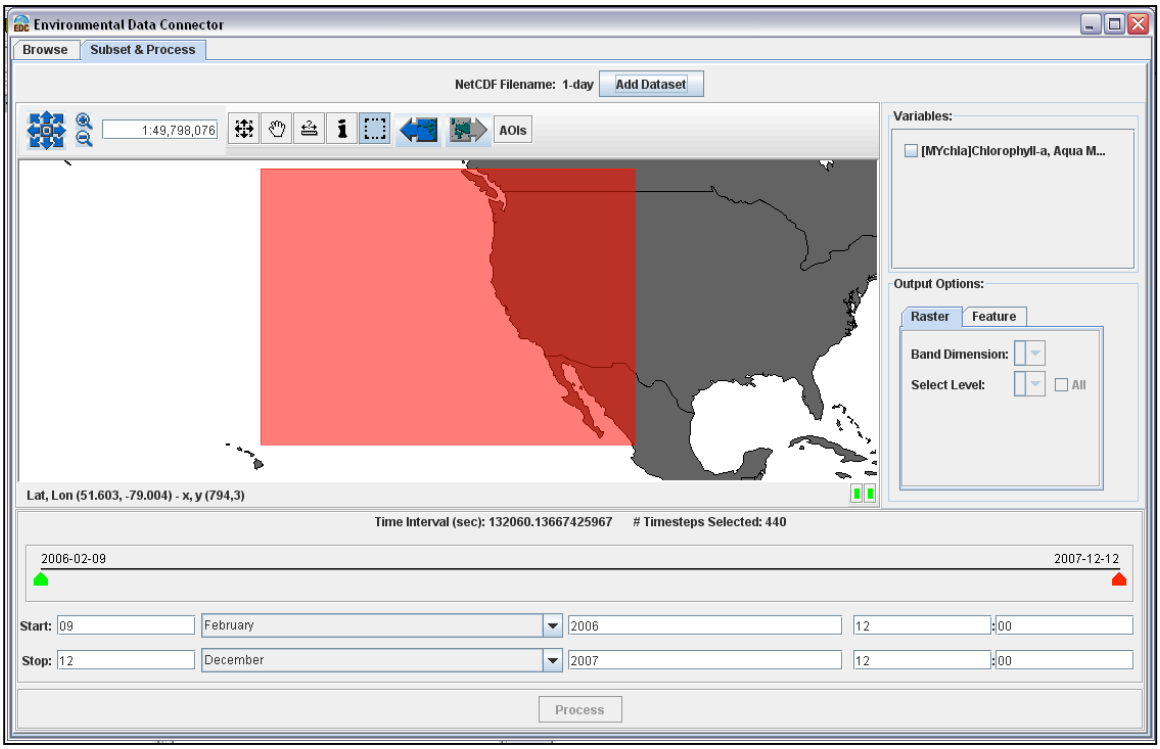


Once you have navigated to a single dataset, you will see that dataset's metadata in the main screen to the right of the directory tree. The amount of metadata will vary depending on the dataset and the data source.

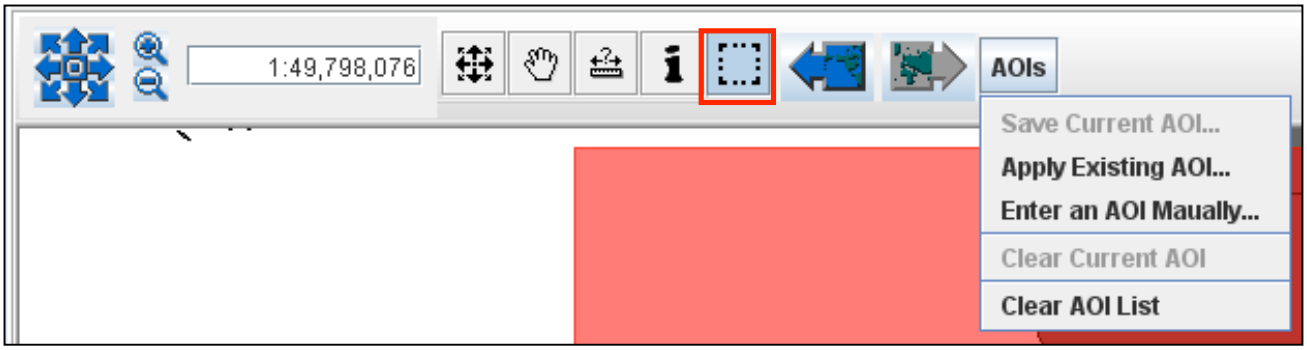
VIEWING AND SUBSETTING DATA

Once you have navigated to your dataset of interest and reviewed the metadata, click the "Subset & Process' button to connect to the dataset. Once the EDC connects to the dataset, a new tab window opens in the EDC. This tab has a 'Subset & Process' title. The 'Browse' tab is still visible, and you may click on this tab to display the catalog tree and connect to other datasets.

In the ‘Subset & Process’ tab, a map is shown with the maximum spatial extent of the dataset shown by a red box.

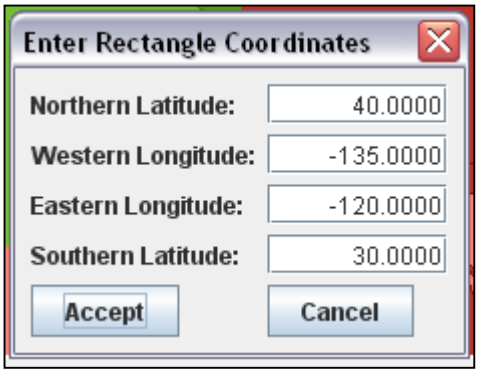


Above the map is a navigation toolbar. This toolbar allows you to navigate around the map. You may zoom in or out and pan around the map. There is also a tool for measuring distances on the map.



SELECTING SPATIAL LIMITS

To select an area, you may either drag a bounding box around the area with the selection tool (outlined above in red), or you may manually input a bounding box using the AOI tool (Area

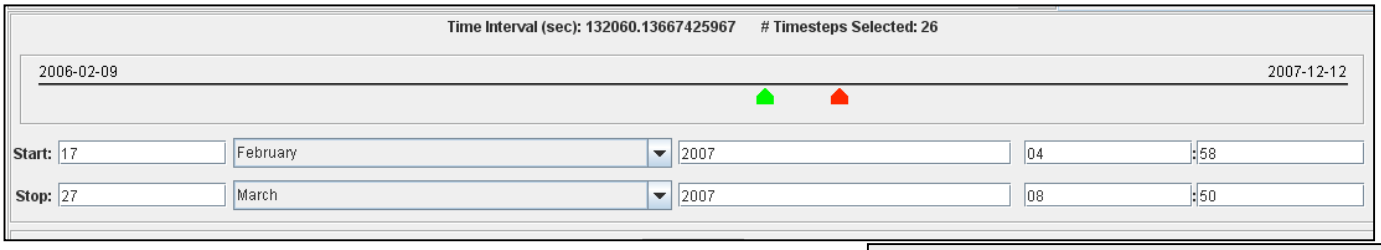


of Interest). With the AOI tool, you may manually input the coordinates of a spatial bounding box for the data. You may also save bounding boxes for future use.

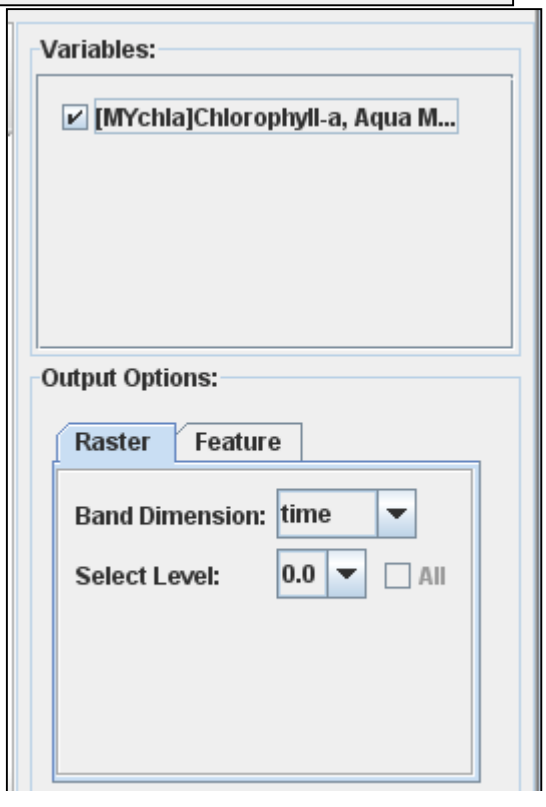
Unfortunately, the current release of the EDC cannot download data which spans the prime meridian (0° longitude). If you input an AOI which spans 0° longitude, only the data which lies to the east (longitude > 0°) will be downloaded. In order to get the entire region of data, you should download the data in two pieces: one piece with west longitude and one piece with east longitude. You may then merge the data in ArcGIS. See the following ESRI technical article on merging multiband rasters: http://support.esri.com/index.cfm?fa=knowledgebase.techarticles.articleShow&d=22005.

SELECTING TEMPORAL LIMITS

Once you have defined the spatial area that you wish to download, define the times of the data that you wish to download. This is done by manipulating the time bar at the bottom of the EDC. In the time bar, the green pointer represents the beginning time of the data to be downloaded, and the red pointer represents the ending time. You can also input start and end times manually (hit the enter button once you've changed the start or end times to update the pointer locations). If you select a beginning time in the time bar that does not exactly coincide with a time in the dataset, then the EDC will download the closest previous timestep, as well as all the data that falls between the beginning and end times.



Just above the time bar, the time interval between consecutive times of the dataset is displayed. This time interval is the average interval between consecutive timesteps for the selected dataset. It is calculated by dividing the range of timesteps (last time – first time) by the number of timesteps. The actual time between two consecutive data points might be different than this value.



DOWNLOADING RASTER DATA

If the data that you are downloading is raster data, then select the 'Raster' tab in the 'Output Options' window in the right hand side of the EDC. Check the box next to the variable that you would like to download.

Depending on the specific dataset you are connected to, there might be more than one variable in this list. If you are downloading the data as a raster, you can only download one variable at a time.

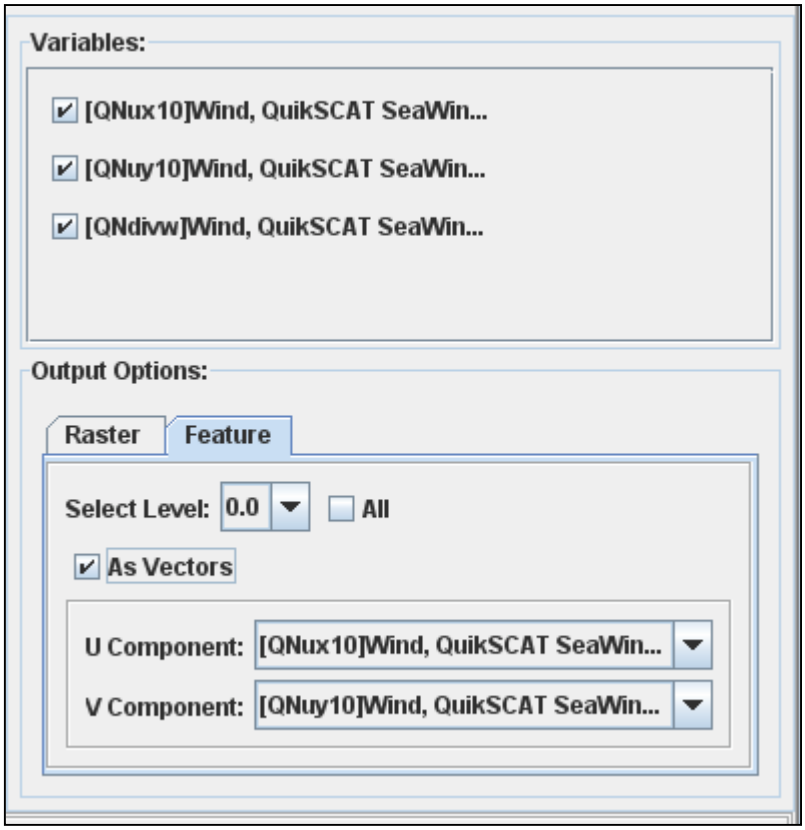
Once you have selected the raster variable, click the 'Process' button to download the data. You will be asked to give a name for the data. Once you click the 'Process' button, the EDC subsets the data on the server, only downloading the data for the time and space that you have selected. The data is initially downloaded in netCDF format. Once the download is complete, the EDC will close. The downloaded netCDF file is then converted to an ArcGIS raster using ArcGIS's conversion tools (only available in ArcGIS version 9.2 or later). Be patient while ArcGIS converts the file. This can often be a lengthy step and there is no progress bar to show that ArcMap is still working on the conversion.

The resulting ArcGIS layer is a multiband raster. The data dimension used as the band dimension is specified in the 'Output Options' window of the EDC. This band dimension is usually time, but can be another dimension (such as depth or height) if that dimension is available. If the dataset contains more than 3 dimensions, then you must select the value of the 4th dimension in the 'Select Level' menu. For example, if the dataset has dimensions of 'latitude', 'longitude', 'altitude', and 'time', and you select 'time' as the band dimension, then you would have to pick the altitude level of the data. Some datasets might have a 4th dimension containing only one value (e.g. – 0 meter altitude for sea level data).

DOWNLOADING FEATURE DATA

If your data is more suited to an ArcGIS feature layer rather than an ArcGIS raster, then you may choose to download the data as a feature dataset. Unlike a raster, a feature dataset can contain multiple data variables. A common type of feature dataset is current or wind data. These datasets consist of (at least) two variables, a u-component (zonal) and a v-component (meridional).

When downloading feature datasets, select the spatial and temporal limits of the data you wish to download. Select the 'Feature' tab in the 'Output Options' window. Select the variables you wish to import in the 'Variables' window.



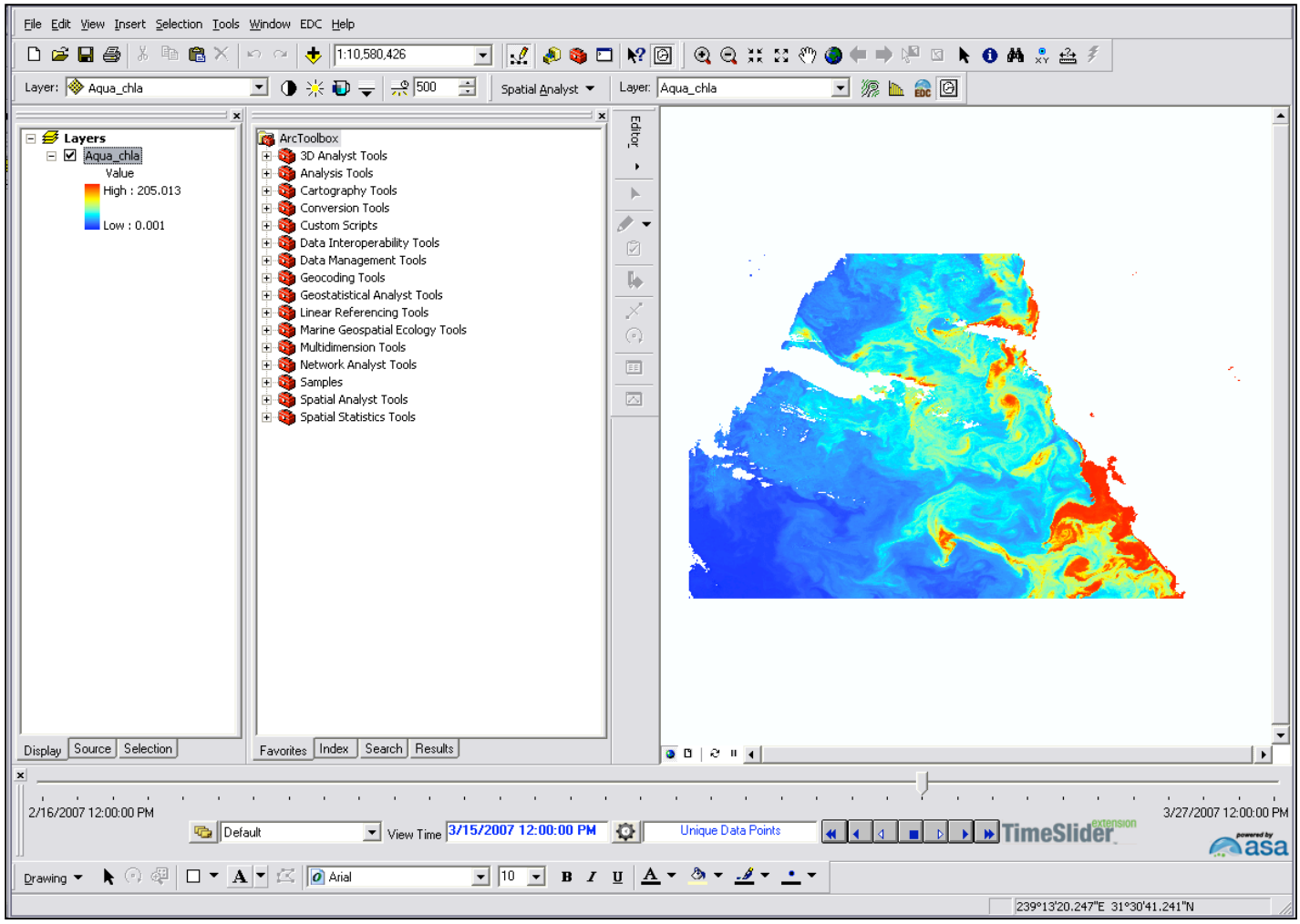
In the 'Feature' tab, select the level of the data you are downloading (see the section on 'Downloading Raster Data' for information on data levels). If the data you want to download is vector data (such as wind or current data), then select the 'As Vectors' box and select the variables which represent the u and v components of the vectors. These variables will be used to calculate the speed and direction fields needed by ArcGIS to display vectors.

Sometimes, you may want to combine variables which are present in different datasets into the same ArcGIS feature dataset. In this situation, subset and process the first dataset you want by connecting to the dataset and selecting the spatial and temporal limits. Then click the "Add Dataset" button, located above the map toolbar. Clicking this button will bring you back to the THREDDS catalog/dataset selection page. Select the second dataset you wish to add, and click "Subset and Process". Both variables will now be visible in the 'Variables' window. Repeat to add additional dataset variables. Note that in order to merge datasets, the metadata for the two datasets being merged must be exactly identical.

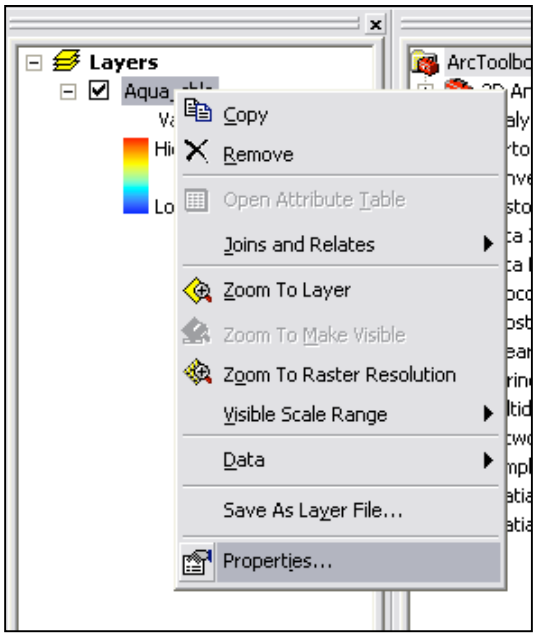
Once you have selected the spatial and temporal limits and selected all the variables you wish to include in the feature dataset, click the 'Process' button to import the data into ArcGIS. The downloading process is similar to the raster case. First, the EDC subsets and downloads the data as a netCDF file. Then, ArcGIS converts this netCDF file to a feature layer. Be patient while ArcGIS converts the netCDF!

VIEWING DATA IN ARCGIS

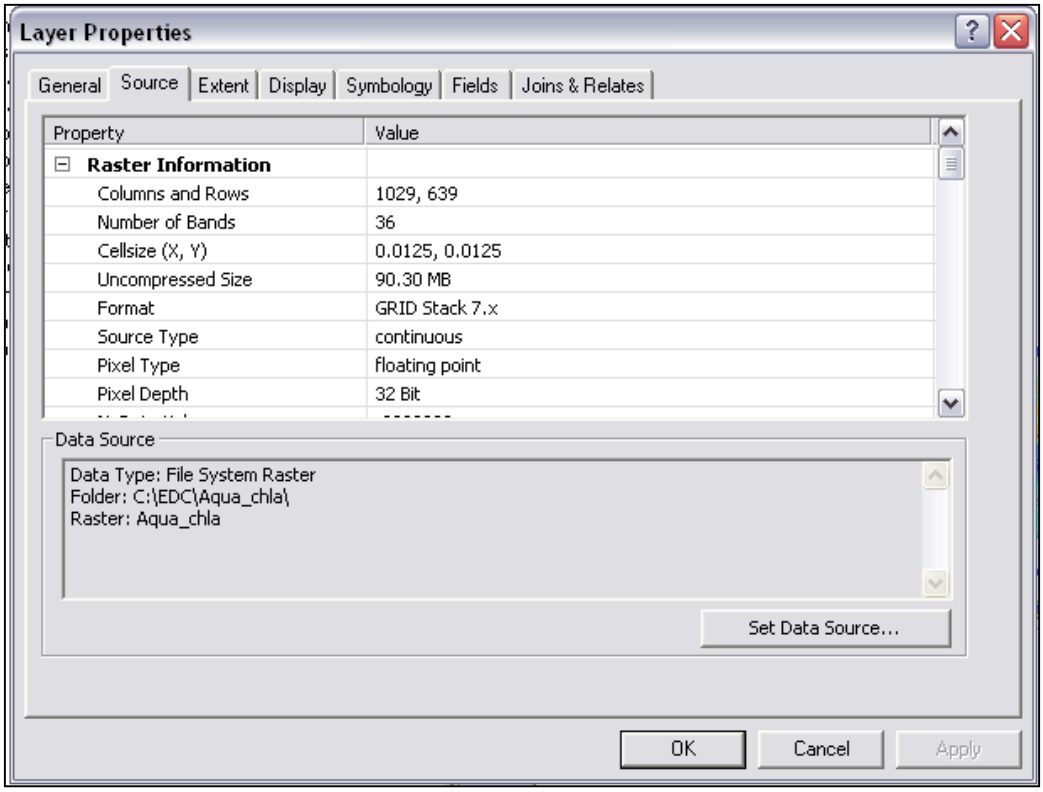
Once the data has been downloaded and converted by ArcGIS tools, it will be displayed in the current ArcMap window.



By right clicking on the dataset in the Table of Contents and selecting 'Properties', you can see the properties of the data that was just imported.



The dataset properties window for a raster datasets shows information such as the source and file location, the spatial extent of the data, the number of bands (the number of bands is equal to the number of timesteps), the cellsize of the raster, and the spatial reference (all data is imported with the WGS1984 spatial reference). In the properties window, you can also change the way that the data is displayed (in the symbology tab).



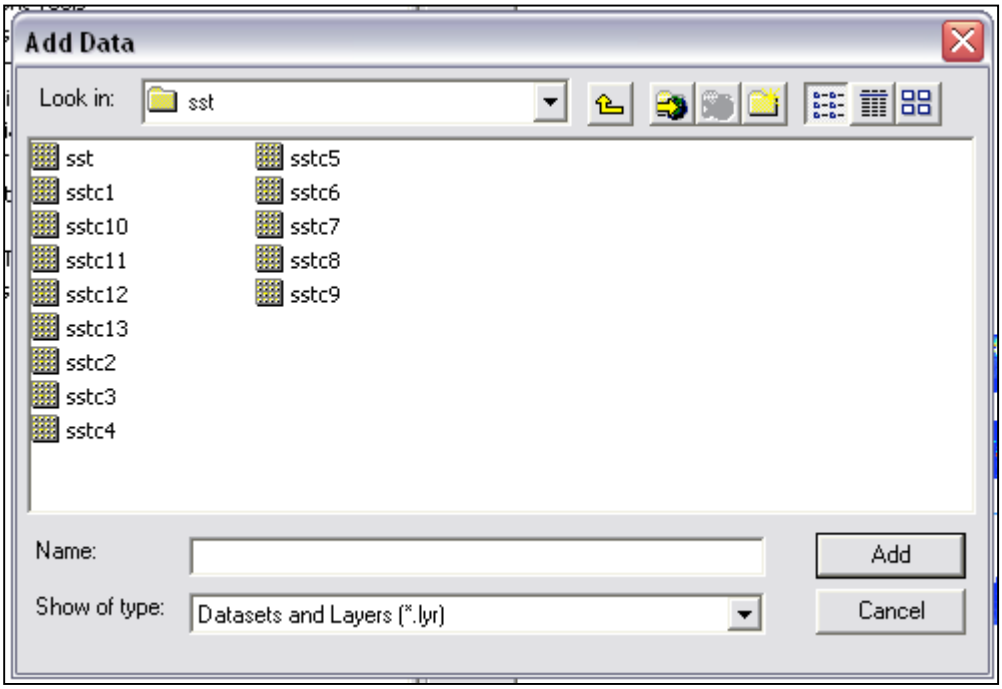
For raster data, the EDC by default sets a linear colormap using the maximum and minimum values for the entire dataset as the upper and lower limits of the colormap. When you animate through the data, the values the colormap does not change as the animation progresses through each raster band. You may change the limits or the colormap in the 'Symbology' tab of the layer properties window.

ADDING PREVIOUSLY SAVED DATA

At some point, you may want to add data which has been previously downloaded. The EDC downloads a netCDF file (with a .nc file extension) and saves this file to a local data directory (the location of this directory can be changed in the file, edcconfig.xml). If this

file is an ArcGIS raster, the netCDF file is then converts this file into a multiband ArcGIS raster. The full multiband raster is named according to the name that you chose when you downloaded the data (e.g., 'sst'). Each band of the raster is also saved as an individual raster named similarly to the full raster with the suffix 'c1', 'c2', 'c3'… attached. These suffixes correspond to the first, second, and third raster bands, respectively. For example, if you had downloaded a data file and called it 'sst', you would have a local directory called 'sst'. In this directory, there would be a netCDF file called 'sst.nc'. There would also exist an ArcGIS raster called 'sst' (the full multiband file), and several files called 'sstc1', 'sstc2', etc., which corresponded to the individual bands.

To import this data into ArcGIS manually, simply add the full multiband raster to ArcMap with the 'Add Data' button (or click the 'File' menu and choose 'Add Data'). If you wish to import the individual bands separately (useful if you plan on doing analysis with those rasters), then import each individual raster with the 'Add Data' button.

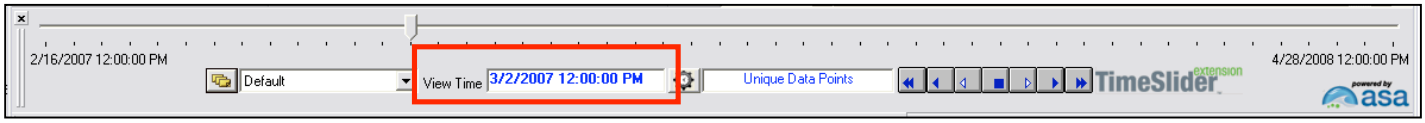


You may also add feature layer data that was downloaded with the EDC. The netCDF file downloaded by the EDC is saved in a folder with the same name as the dataset. This netCDF file is then converted into feature classes which are stored in a personal geodatabase with the same name as the dataset. Inside this personal geodatabase, there is a feature class for each variable that was downloaded (named for the variable, not the dataset), as well as a feature class which is the join of all the downloaded variables. This joined feature class is named the same as the dataset, with the suffix "_Join" added. For example, if you downloaded a vector dataset consisting of variables "x" and "y" and called it "wind", then the EDC will create a local directory called "wind". Inside this directory will be the netCDF file called "wind.nc". There will also be a personal

geodatabase called “wind.mdb”. Inside this personal geodatabase, there will be feature classes called “x”, “y”, and “wind_Join”.

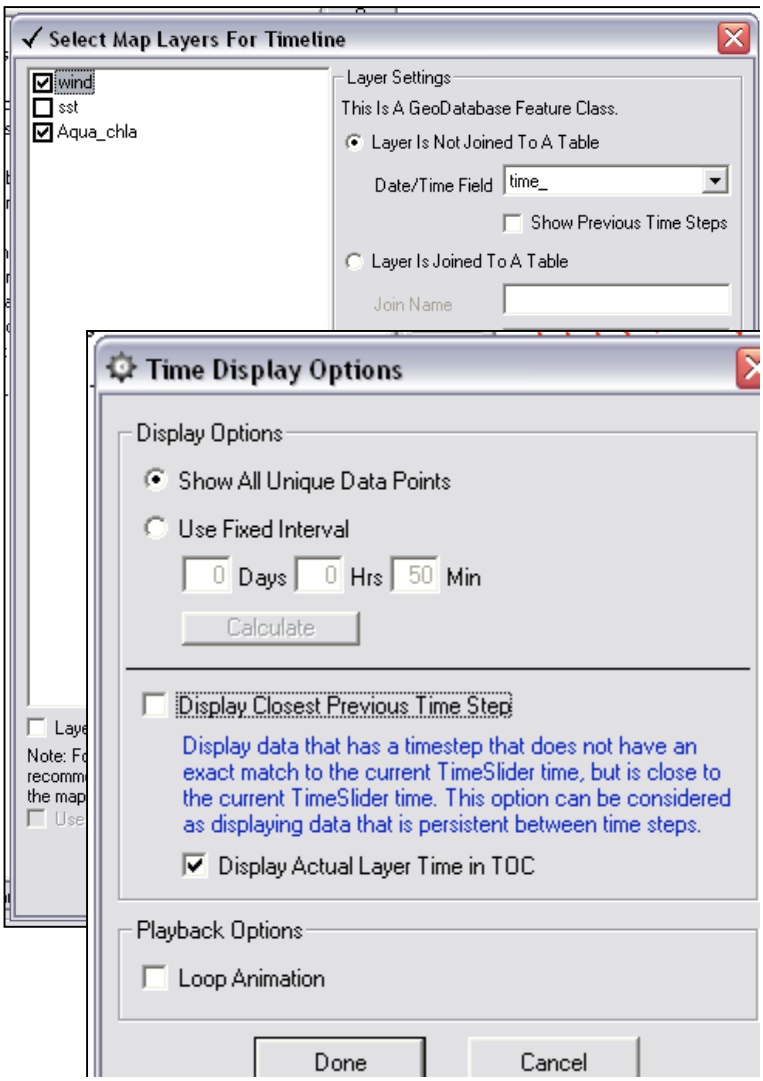
ANIMATING DATA IN ARCGIS

Once you import the data into ArcGIS, you will see the times of the data appear in the TimeSlider toolbar. The different time steps that were imported are indicated by tick marks on the TimeSlider toolbar at the bottom of the screen. You may slide the time marker around manually to change the displayed time, or you may use the TimeSlider tools to animate the data. The tools allow you to step through the time steps forward or backward in time, or to animate through the data forward to backward in time. The current time being displayed is shown in the “View Time” window (outlined below in red)



FEATURES AND BUTTONS OF THE TIMESLIDER TOOLBAR

“Select Map Layers” Button – This window allows you to select which map layers will be animated by the TimeSlider. If the layer is a feature layer, then you need to specify the field which contains time information. This field must contain time information in string format. If the data has been imported using the EDC, then the field is called “time_”. You can also specify the speed and direction fields if the feature layer contains vector data. In this case, check the “Layer Contains Vectors” box and specify which fields contain the speed and direction information.



The drop down menu next to the “Select Map Layers” button allows you to choose which data’s time steps are shown along the TimeSlider’s time bar.

“Time Display Options” Button –

This window allows you to change how the TimeSlider animates data. If your data is on a fixed interval (e.g. one day apart) with no missing data points, you can select the 'Use Fixed Interval' Option. Be sure that the interval shown is correct. If your data is missing some data points or if there is unequal intervals between data points (e.g. monthly data), use the "Show All Unique Data Points" option.

If the "Display Closest Previous Time Step" box is checked, the TimeSlider will display data even if it does not have a time exactly equal to the current TimeSlider time. In this case, the TimeSlider will display the closest previous time step.

If the "Display Actual Layer Time in TOC" box is checked, then the currently displayed time will be shown alongside the layer name in ArcMap's Table of Contents.

If the "Loop Animation" box is checked, then the TimeSlider will keep animating data, looping back to the first timestep once the last timestep has been displayed, until the animation is stopped manually.

More information on the TimeSlider Extension for ArcGIS is available at ASA's TimeSlider Extension website: http://www.asascience.com/TimeSlider/index.htm.

RUNNING THE EDC OUTSIDE OF ARCGIS (STANDALONE MODE)

It is possible to run the Environmental Data Connector outside of ArcGIS. In this case, the EDC subsets and downloads data as a netCDF file. The data is stored on disc as a netCDF and is not converted to ArcGIS format. You may work with data in netCDF format, or convert it to ArcGIS format using ArcGIS tools. The standalone version of the EDC requires Java 1.5. ArcGIS is not needed.

There is an archive titled 'EDC_Standalone' which contains the java files necessary to run the EDC. Once you have downloaded the archive, unzip the archive to any folder on your computer. In MS Windows, double-click the file "launchEDC.bat" to run the EDC. To run the standalone EDC on a Mac, open a terminal window and change the directory to the EDC_Standalone directory. Type the command "./launchEDC.sh" to launch the EDC.

CONFIGURING THE EDC

Certain parts of the EDC may be configured according to the user's preference. These parameters are contained in the 'edcconfig.xml' file, which is contained in the

'EDC/System' folder. Open this file in a text editor to change these parameters (WordPad in MS Windows is a good text editor).

Each parameter is contained in a text block. Each text block contains information about the parameter the values the parameter can take on. For example, the text block for the parameter 'CLOSE_AFTER_PROCESSING' looks like:

<!--CLOSE_AFTER_PROCESSING: MUST BE A BOOLEAN VALUE.
If true, the application will close after processing a dataset.
If false, the application will remain open until it is closed maually.
-->
<CLOSE_AFTER_PROCESSING>*true*</CLOSE_AFTER_PROCESSING>

This parameter must be a Boolean value ('true' or 'false'). To edit this parameter, change the text between the <CLOSE_AFTER_PROCESSING> tags (shown in italics) to read either 'true' or 'false'.

Some important parameters that you might need to edit are: (not an exhaustive list)

DISPLAY_TYPE: Changes the look and processing steps of the EDC. The "General" (0) display type is the most simple, and is best suited to running the EDC in standalone mode. In this display type, the EDC only downloads data in netCDF format. No additional processing is done. The "ESRI" (1) display type is most suited to using the EDC is conjunction with ArcGIS. In this display type, the EDC will download data and then convert the data to ArcGIS format.

CLOSE_AFTER_PROCESSING: Must be true when using the EDC with ArcGIS. If this value is not true, then the EDC will remain open after downloading data and ArcGIS will not convert and import the downloaded files. If running the EDC in standalone mode, then this can be set to false in order to download multiple files in one EDC session.

USE_SUBDIRECTORIES: If true, then a subdirectory with the same name as the dataset will be created in the default output directory. All dataset files will be put in this created directory. If false, then all data will be placed directly in the default output directory.

OUTPUT_LOCATION: A string value indicating the directory location of all downloaded data. If this field is empty, then all data will be placed in the EDC\ directory.

ALLOW_FILE_REPLACEMENT: If true, then the user can overwrite existing files. If false, then the EDC will not allow file overwriting. Note that if the 'DISPLAY_TYPE' field is set to ESRI, then this value will always be false. Because of the file access restrictions that ESRI places on certain files, the EDC cannot overwrite files once they have been opened in ArcMap, even if the file is no longer in use.

USE_VARIABLE_NAME_FOR_OUTPUT: This controls the default name EDC gives to data. If this is true, then the name of the first variable selected will be used as the default output name. If this is false, then the description (the 'long_name') will be used as the default. The long name is a more descriptive name for the dataset, but sometimes is too long to be used as a file name in ArcGIS.

HEAP_SIZE: The amount of memory (in megabytes) allocated for the EDC to run. This is approximately equal to the maximum amount of data which can be downloaded. Note that there is currently no way to tell whether you will exceed this limit before you begin downloading data. You may change this value to be as large as you wish, but download times may become *very* long.